



OPEN

Bayesian Inference of Natural Rankings
in Incomplete Competition NetworksJuyong Park¹ & Soon-Hyung Yook²SUBJECT AREAS:
INFORMATION THEORY
AND COMPUTATION
APPLIED MATHEMATICSReceived
30 July 2013Accepted
5 August 2014Published
28 August 2014Correspondence and
requests for materials
should be addressed to
J.P. (juyongp@kaist.ac.
kr)

¹Graduate School of Culture Technology, Korea Advanced Institute of Science and Technology, Daejeon, Republic of Korea 305-701 and Physics Department, Kyung Hee University, Seoul, Republic of Korea 130-701, ²Physics Department, Kyung Hee University, Seoul, Republic of Korea 130-701.

Competition between a complex system's constituents and a corresponding reward mechanism based on it have profound influence on the functioning, stability, and evolution of the system. But determining the dominance hierarchy or ranking among the constituent parts from the strongest to the weakest – essential in determining reward and penalty – is frequently an ambiguous task due to the incomplete (partially filled) nature of competition networks. Here we introduce the “Natural Ranking,” an unambiguous ranking method applicable to a round robin tournament, and formulate an analytical model based on the Bayesian formula for inferring the expected mean and error of the natural ranking of nodes from an incomplete network. We investigate its potential and uses in resolving important issues of ranking by applying it to real-world competition networks.

Understanding of the structure and dynamics of complex networks found in nature, society, and elsewhere have been greatly facilitated by remarkable advances in the modern science of networks^{1,2}. Fundamental network problems that have garnered interest among natural scientists include the highly skewed (approximately power-law) degree (connectivity) distributions, identification of communities or modules, and various critical phenomena and their implications on the functioning and stability of networked systems^{3,5,6}. “Centrality”, the measure of the importance or superiority of a node in the network, is another concept very often studied^{7,8}. The best known modern example is Google's PageRank⁹, whose algorithm for calculation can be understood as a random walk (surfing) problem along the hyperedges connecting the web pages on the Worldwide web.

The idea of ranking the nodes based on their relative strengths or relevance can be useful for understanding the underlying dynamics of many networks; in fact, in many complex systems – natural, social, or man-made – the competition-and-reward mechanism is an essential ingredient of their functioning and evolutionary dynamics. The dominance hierarchy or ranking refers to the linear ordering of things from the strongest to the weakest based on the results of competitions or comparisons¹⁰. In the case where the things undergo pairwise (one-to-one) competitions, the entire set of competitions can be represented as a directed network where an arrow points from the winner to the loser of a competition (see Fig. 1 (a) and (b)) or vice versa, depending on convention. Food webs (of predators and preys) in ecological systems, the domination-submission interaction networks of animals (observed in birds, American bison, etc.¹⁰), sport schedules, and certain types of elections or voting systems where the candidates are compared pairwise (such as the Condorcet method, to be discussed later) are widely-known examples of competition networks. The dominance hierarchy may assume different names depending on the domain – “trophic levels” in ecology, and “ranking” or “standings” in sports, for example. In the remainder of this letter, for convenience we shall often utilize sports terminology, e.g. ranking, contestant (or player or team), game, match, win, loss, tie, and so forth. Among competition networks a tournament is one in which every player competes against everybody else, also called a round robin. It can be represented as a complete (full network) with a directed edge between every pair of nodes, so that in this paper we also call one a complete competition. In such as case, determining the ranking is straightforward: We can simply rank the players in the decreasing order of their total wins, i.e. the out-degree k^{out} . When there exists a tie in k^{out} , we can employ the following “tie breaker”: We consider the reduced round robin composed of those that are tied, and rank them according to their wins therein. This can be applied iteratively to continue breaking the ties that persist. (Note that no tie may be further broken in some cases, for instance when three teams i, j , and l have the same total wins ($k_i^{\text{out}} = k_j^{\text{out}} = k_l^{\text{out}}$) and i lost to j , j lost to l , and l lost to i , i.e. $\{\sigma_{ij}, \sigma_{jl}, \sigma_{li}\} = \{1, 1, 1\}$ in the adjacency matrix notation. In such a case we may adopt yet another tie breaker such as the total points scored in games). We call the ranking of nodes obtained this

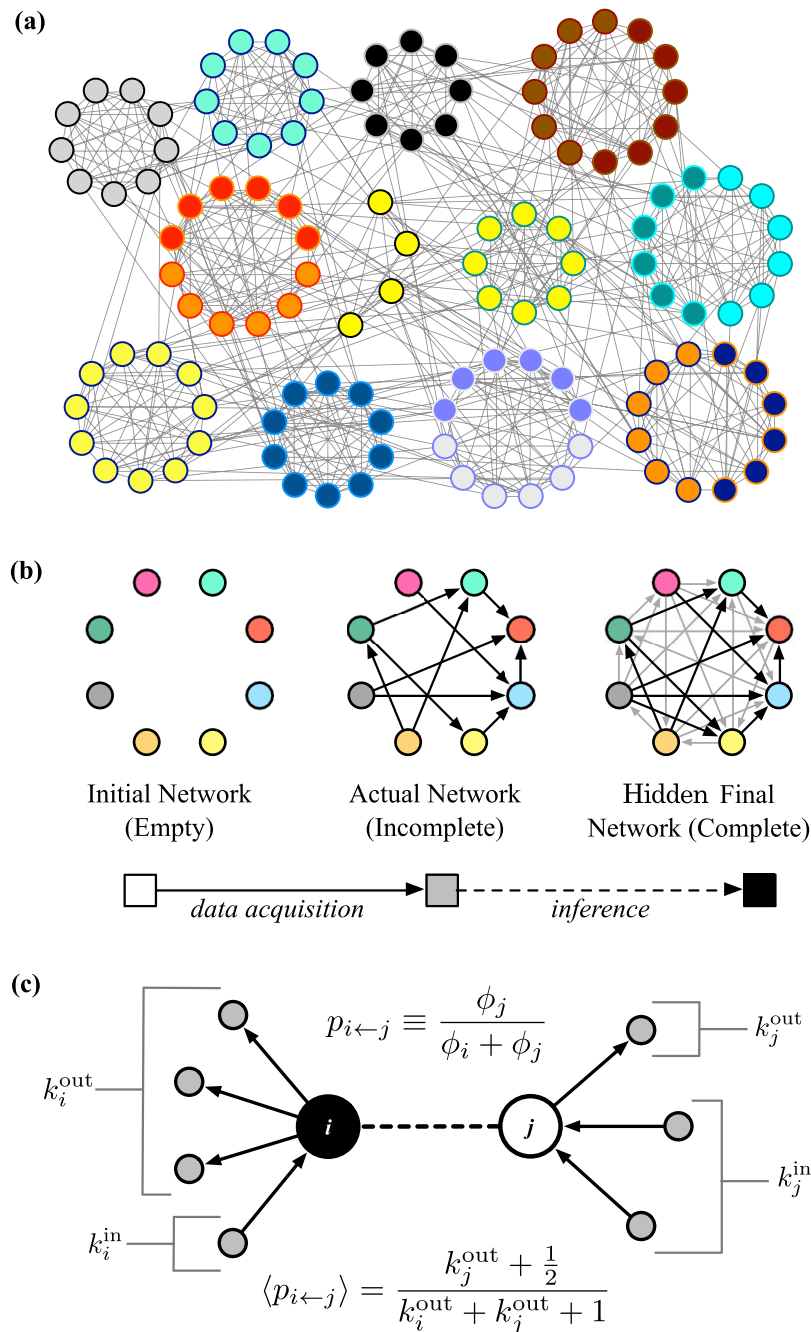


Figure 1 | (a) The incomplete competition schedule network of US college football. The nodes represent universities, and edges represent games. The network is incomplete (partially filled), with its connectance — the fraction of node pairs that are connected — typically $\sim 10\%$. This causes uncertainties in the rankings determined based on the outcomes of the games. The result of a game can be represented as a directed edge pointing from the winner to the loser. (b) The incomplete competition network (middle) can be thought of as an intermediary stage of a schedule that starts as empty (left) and ends as complete, when all possible games have been played (right). The progression of the season is the data acquisition process that forms the basis for inference of the expected natural ranking. (c) The setup of a single strength parameter (ϕ) model for estimating the expected win probability of a team in a potential contest.

way the “Natural Ranking,” as it results from a tournament and is the fairest – every player competes against every other. Note that this is applicable to multiple round-robins as well, as long as each node pair plays an equal number of times.

We note that the earliest recorded form of this scheme is by Ramon Llull⁴, and is now also called the Condorcet method for voting (on bills in the parliament, for instance). Lull’s original formulation is analogous to the round-robin in which after a full round of voting between two candidates (also called alternatives) the one

that has won the most pairings is chosen. A more common formulation of the Condorcet system is one in which each alternative is given an order of preference by the voters (although the voters are not obligated to give every alternative a preference), and the winner of a pairwise comparison is one that has been preferred by more voters. Ranking and voting are intimately related, and much effort has been put into understanding the issues as well¹¹.

Despite its simplicity and intuitiveness, the natural ranking introduced here is often inapplicable to many real-world networks, as they



are often incomplete. Expecting a real-world competition to be complete is perhaps unreasonable in practice for several reasons. First, the cost of a complete competition can be very high even for moderately large systems: In a network of n contestants, the minimum number of competitions required is $\binom{n}{2} \simeq n^2/2$. For the case of the popular US college football of 120 teams for instance (Fig. 1 (b)) there may simply be not enough number of weeks for each team to play the 119 games necessary in a year for those who care for the athletes' health (or their education). Second, there may be insurmountable physical constraints as in an ecological food web where the spatial separation between the habitats of two species may hinder them from interacting directly¹². In this paper we propose an analytical method to infer (estimate) the natural rankings from an incomplete network.

Results

The final natural ranking from an incomplete network can be estimated by considering the actual (incomplete) competition network as an intermediate stage of the "schedule" of a complete competition, between an empty network and a hidden complete network where all competitions have been made (see Fig. 1 (c)). This then becomes the problem of inferring the future of the system of variables based on current information (i.e. data), for which the Bayesian one of the most widely used frameworks¹³. The Bayesian framework is a process in which the estimate (called the prior) of the distribution $\pi(x)$ of parameter x is updated into a new estimate (called the posterior) $\pi(x|D)$ when new data D is made available, via the following Bayes' formula^{13,14}:

$$\pi(x|D) \propto \mathcal{L}(D|x) \times \pi(x), \quad (1)$$

where $\mathcal{L}(D|x)$, called the likelihood, is the probability of D being observed given a parameter value x . At the next round of update with new available data, the posterior becomes the prior. (This is also called Recursive Bayes).

Here we use the Bayes' formula Eq. (1) to estimate $\{\tilde{k}^{\text{out}}\}$, the projected total wins (outdegree) based on an incomplete competition network to infer the natural ranking. \tilde{k}_i^{out} of node i is the sum of two quantities: The number of actual wins thus far, \tilde{k}_i^{out} , and the expected number of wins from unplayed games, that is the sum of the probabilities of winning the games. Thus our goal becomes estimating $p_{j \leftarrow i} = p_{ji}$ given the current state of the competition. We find p_{ji} consistent with Eq. (1) via the following steps. First, when we have no basis on which to judge the two teams' strengths, for examples at the beginning of the season when the teams have not played any games, we are maximally ignorant or uncertain of p_{ji} . A naïve guess would be a flat (uniform) prior $\pi(p_{ji}) = 1$ where all p_{ji} values are equiprobable. A more common choice for this type of a problem is the Jeffreys prior given as

$$\pi(p_{ij}) = \frac{\Gamma(1)}{\Gamma(1/2)\Gamma(1/2)} p_{ij}^{-1/2} (1-p_{ij})^{-1/2}, \quad (2)$$

due to the desirable property that it is invariant under re-parametrization of the parameter (p_{ij}) thanks to being proportional to the Fisher information. The idea behind Jeffreys prior and its history of development is an interesting topic in itself, which are discussed more deeply in Ref. 14, 15.

Now assume that we observe that i loses to j , i.e. $D = \{\sigma_{ji} = 1\}$, whose probability is simply p_{ji} . From Eqs. (1) and (2) we obtain

$$\pi(p_{ij}|D) = \frac{\Gamma(2)}{\Gamma(3/2)\Gamma(1/2)} p_{ij}^{1/2} (1-p_{ij})^{-1/2}. \quad (3)$$

The above procedure will serve as the foundation for our method of estimating the $\pi(p_{ij})$ for a node pair that is yet to play. We introduce a

strength parameter $\phi_i \in [0, \infty)$ for each contestant such that p_{ij} between two contestants is

$$p_{ij} \equiv \frac{\phi_j}{\phi_i + \phi_j}. \quad (4)$$

In terms of $\pi(\phi_i)$, the distribution of ϕ_i , we can now write $\pi(p_{ij})$ as

$$\pi(p_{ij}) \equiv \int_0^\infty \int_0^\infty \delta\left(p_{ij} - \frac{\phi_j}{\phi_i + \phi_j}\right) \pi(\phi_i) \pi(\phi_j) d\phi_i d\phi_j. \quad (5)$$

An essential step in our formalism is finding the exact form of $\pi(\phi)$ that renders this expression consistent with Bayes' formula, Eq. (1). For the initial Jeffreys prior, Eq. (2), for instance, we find $\pi(\phi) = \phi^{-1/2} e^{-\phi} / \Gamma(1/2)$ for both i and j gives the correct Bayes' formula. In the case of Eq. (3) (i.e. $D_1 = \{\sigma_{ij} = 1\}$) we see that the following change is appropriate:

$$\begin{aligned} \pi(\phi_i|D_1) &= \pi(\phi_i) \quad (\text{unchanged}) \\ \pi(\phi_j|D_1) &= \frac{\phi_j^{1/2} e^{-\phi_j}}{\Gamma(3/2)}. \end{aligned} \quad (6)$$

This agrees with our intuition that $\sigma_{ij} = 1$ means that i is likely to be weaker than j , since using these posteriors we find $\langle \phi_i \rangle = \frac{1}{2}$ and $\langle \phi_j \rangle = \frac{3}{2}$.

This procedure can be repeated to find a general pattern. Assume that now j , having defeated i , competes against l with $\pi(\phi_l) = \phi_l^{-1/2} e^{-\phi_l} / \Gamma(1/2)$. Then, using Eqs. (5) and (6) we find $\pi(p_{ij}) \propto p_{kj}^{1/2} (1-p_{lj})^{-1/2}$ between j and l . Using the Bayes' formula, the possible posteriors are

$$\pi(p_{ij}|D_2) \propto \begin{cases} p_{ij}^{3/2} (1-p_{ij})^{-1/2} & \text{if } D_2 = \{\sigma_{lj} = 1\} \\ p_{ij}^{1/2} (1-p_{ij})^{1/2} & \text{if } D_2 = \{\sigma_{jl} = 1\}, \end{cases} \quad (7)$$

depending on D_2 , the outcome of the game between l and j . It turns out that the following update rules for the *winner's* strength are consistent with the Bayes' formula (again, no changes are necessary for the loser's):

$$\begin{aligned} \pi(\phi_j|D_2 D_1) &= \frac{\phi_j^{3/2} e^{-\phi_j}}{\Gamma(5/2)} \quad \text{if } \sigma_{lj} = 1 \\ \pi(\phi_l|D_2) &= \frac{\phi_l^{1/2} e^{-\phi_l}}{\Gamma(3/2)} \quad \text{if } \sigma_{jl} = 1. \end{aligned} \quad (8)$$

In general, the strength distribution of a contestant dependent on its accumulated wins k^{out} in the form

$$\pi(\phi|k^{\text{out}}) = \frac{\phi^{k^{\text{out}} - 1/2} e^{-\phi}}{\Gamma(k^{\text{out}} + 1/2)} \quad (9)$$

leads correctly to the following estimate between two teams with k_i^{out} and k_j^{out} wins:

$$\pi(p_{ij}|k_i^{\text{out}}, k_j^{\text{out}}) = \frac{\Gamma(k_i^{\text{out}} + k_j^{\text{out}} + 1)}{\Gamma(k_i^{\text{out}} + 1/2) \Gamma(k_j^{\text{out}} + 1/2)} (1-p_{ij})^{k_i^{\text{out}} - 1/2} p_{ij}^{k_j^{\text{out}} - 1/2} \quad (10)$$

With these we can now calculate $\langle p_{ji} \rangle$, the expected win score (out-degree) gain of i against j :



$$\begin{aligned}\langle p_{ji} \rangle &\equiv \int_0^1 p_{ji} \times \pi(p_{ji} | k_i^{\text{out}}, k_j^{\text{out}}) dp_{ji} \\ &= \int_0^\infty \int_0^\infty \frac{\phi_i}{\phi_i + \phi_j} \times \pi(\phi_i | k_i^{\text{out}}) \pi(\phi_j | k_j^{\text{out}}) d\phi_i d\phi_j = \frac{k_i^{\text{out}} + \frac{1}{2}}{k_i^{\text{out}} + k_j^{\text{out}} + 1}.\end{aligned}\quad (11)$$

At any given point in the competition, therefore, the expected final outdegree W_i for team i is

$$\tilde{k}_i^{\text{out}} = k_i^{\text{out}} + \sum_{j \in \Omega_i} \langle p_{ji} \rangle = k_i^{\text{out}} + \sum_{j \in \Omega_i} \frac{k_i^{\text{out}} + \frac{1}{2}}{k_i^{\text{out}} + k_j^{\text{out}} + 1}, \quad (12)$$

where Ω_i is the set of nodes that i is yet to play against. In a complete competition network the second term vanishes, while in an incomplete network it differentiates two teams with the same k^{out} . From its functional form we can also tell that having beaten stronger opponents gives one an advantage — the unplayed opponent's k^{out} appears in the denominator — thereby naturally incorporating the “strength of schedule.”

We can also compute the expected variance of the final outdegree given as

$$\begin{aligned}\Delta^2 \tilde{k}_i^{\text{out}} &= \left\langle \left(\tilde{k}_i^{\text{out}} - \langle \tilde{k}_i^{\text{out}} \rangle \right)^2 \right\rangle \\ &= \sum_{j \in \Omega_i} \langle p_{ji} \rangle (1 - \langle p_{ji} \rangle) + 2 \sum_{(j < l) \in \Omega_i} (\langle \sigma_{ji} \sigma_{li} \rangle - \langle \sigma_{ji} \rangle \langle \sigma_{li} \rangle) \\ &= \sum_{j \in \Omega_i} \frac{(k_i^{\text{out}} + \frac{1}{2})(k_j^{\text{out}} + \frac{1}{2})}{(k_i^{\text{out}} + k_j^{\text{out}} + 1)^2} + 2 \sum_{(j < l) \in \Omega_i} (\langle \sigma_{ji} \sigma_{li} \rangle - \langle \sigma_{ji} \rangle \langle \sigma_{li} \rangle).\end{aligned}\quad (13)$$

We note that the second term is non-vanishing, since $\langle \sigma_{ji} \sigma_{li} \rangle \neq \langle \sigma_{ji} \rangle \langle \sigma_{li} \rangle$ due to the shared index i ; we call $\sigma_{ji} \sigma_{li}$ *connected*, analogous to connected Feynman diagrams in quantum field theory, which have also found some uses in network theory^{16,17}. $\langle \sigma_{ji} \sigma_{li} \rangle$ is

$$\begin{aligned}\langle \sigma_{ji} \sigma_{li} \rangle &= \int_0^\infty \int_0^\infty \int_0^\infty \frac{\phi_i}{\phi_i + \phi_j} \frac{\phi_i}{\phi_i + \phi_l} \times \pi(\phi_i | k_i^{\text{out}}) \pi(\phi_j | k_j^{\text{out}}) \pi(\phi_l | k_l^{\text{out}}) d\phi_i d\phi_j d\phi_l \\ &= \frac{1}{\Gamma(k_i^{\text{out}} + \frac{1}{2})} \int_0^\infty \phi_i^{k_i^{\text{out}} + k_j^{\text{out}} + k_l^{\text{out}} + \frac{1}{2}} \Gamma\left(-k_j^{\text{out}} + \frac{1}{2}, \phi_i\right) \Gamma\left(-k_l^{\text{out}} + \frac{1}{2}, \phi_i\right) e^{-\phi_i} d\phi_i \\ &\equiv \mathcal{B}\left(k_i^{\text{out}}, k_j^{\text{out}}, k_l^{\text{out}}\right),\end{aligned}\quad (14)$$

where $\Gamma(a, z) \equiv \int_z^\infty t^{a-1} e^{-t} dt$ is the incomplete gamma function.

No closed solution for $\mathcal{B}(k_i^{\text{out}}, k_j^{\text{out}}, k_k^{\text{out}})$ exists at the time of this writing, so we resort to numerical evaluation. Finally, the variance is

$$\begin{aligned}\Delta^2 \tilde{k}_i^{\text{out}} &= \left(k_i^{\text{out}} + \frac{1}{2}\right) \sum_{j \in \Omega_i} \frac{k_j^{\text{out}} + \frac{1}{2}}{(k_i^{\text{out}} + k_j^{\text{out}} + 1)^2} \\ &\quad + 2 \sum_{(j < l) \in \Omega_i} \left[\mathcal{B}\left(k_i^{\text{out}}, k_j^{\text{out}}, k_l^{\text{out}}\right) - \frac{(k_i^{\text{out}} + \frac{1}{2})^2}{(k_i^{\text{out}} + k_j^{\text{out}} + 1)(k_i^{\text{out}} + k_l^{\text{out}} + 1)} \right].\end{aligned}\quad (15)$$

Application to Real Competition Networks

We now showcase our method by applying to two well-known competition networks that feature unique challenges, namely American college football and the English Premier League soccer.

American college football: Incomplete competitions and the number of playoff rounds. The American college football network is incomplete, with each university playing against merely $\sim 10\%$ of the nodes in each season. The popularity of the sport and the desire

for an annual national championship — “Who is the best?” is perhaps sports fans’ biggest interest — has resulted in the invention of ranking systems that are purported to overcome the deficiency. The popularity of the sport and the substantial benefits, financial and otherwise, awarded to the champions render a robust and fair ranking method essential. The official ranking system called Bowl Championship Series (BCS) used until 2013 combined human polls and mathematical formulae to select determine the two “best” teams for annual the national championship^{18–20}. This is to be succeeded by College Football Playoff (CFP for short) system where four teams will contend in a two-round playoff series starting in 2014. The increase from two to four is intended to overcome the criticism of the BCS system by dissatisfied fans who argued that choosing only two teams from a pool of more than a hundred is insufficient given the low connectance of $\sim 10\%$ of the schedule network. While it may be too early to truly assess the efficacy of the CFP system, issues raised by the sparsity of a schedule network are worth exploring in their own right, which we tackle with our method.

The result of our method applied to American college football of 2010 is shown in Fig. 2 (a). It shows the estimated final outdegrees $\{\tilde{k}^{\text{out}}\}$ with the error bars indicating the squared-root-variance $\left\{ \left(\Delta^2 \tilde{k}^{\text{out}} \right)^{1/2} \right\}$ as the measure of the uncertainty. As expected, teams with the same number of actual wins k^{out} are separated by the strength-of-schedule term in Eq. (11). The $\left(\Delta^2 \tilde{k}^{\text{out}} \right)^{1/2}$ further allows an interesting interpretation: Taking $\tilde{k}^{\text{out}} \pm \left(\Delta^2 \tilde{k}^{\text{out}} \right)^{1/2}$ as the reasonable expected range of the final outdegree, Fig. 2 (a) implies that the proposed four-team playoff system may still be quite insufficient: the first-ranked team (University of Texas at Austin) has an overlapping win score range with those of teams up to the 32nd-ranked team (University of Nevada), suggesting that a further expansion of the playoff system to include 32 teams would not be unreasonable.

As the connectance of the network increases the uncertainty is bound to decrease, so it would be interesting to see how the score range overlap changes as a function of the increasing connectance. In Fig. 2 (b) we show the number of teams with an overlapping score ranges with the first-ranked team. Beyond the actual number of games (679 games), we generated 1 000 simulated complete seasons based on the final fitness. Our method indicates that on average a connectance of 0.29 would be necessary for a 16-team playoff, 0.54 for an 8-team playoff, etc.

English premier league football: complete network and stabilization of rankings. We now apply our method to the English Premier League (EPL) soccer network. The EPL network presents another interesting opportunity for our method. The EPL network, composed of 20 teams, is complete with every pair of teams playing twice in a given season. As the true natural ranking will be revealed in a complete network, the question of the sufficient size of a playoff system is not relevant here, unlike American college football.

One of the interesting questions in such a complete network is the convergence of the teams’ rankings. In Fig. 3 (a) we show the \tilde{k}^{out} of teams chosen from three distinct tiers, Manchester City from the top tier (with final score 30.5, blue), Liverpool from the middle tier (final score 19, red), and Wolverhampton from the bottom tier (final score 10, green), as the function of games played in the 2011–2012 season. We observe significant fluctuations in the standings in the beginning that attenuate and to reveal clear standing as the season progresses. In Fig. 3 (b) we show the number of teams with an overlapping \tilde{k}^{out} range with the first-ranked team, which reaches a minimum of 2 (Manchester City and Manchester United ended up tying with the

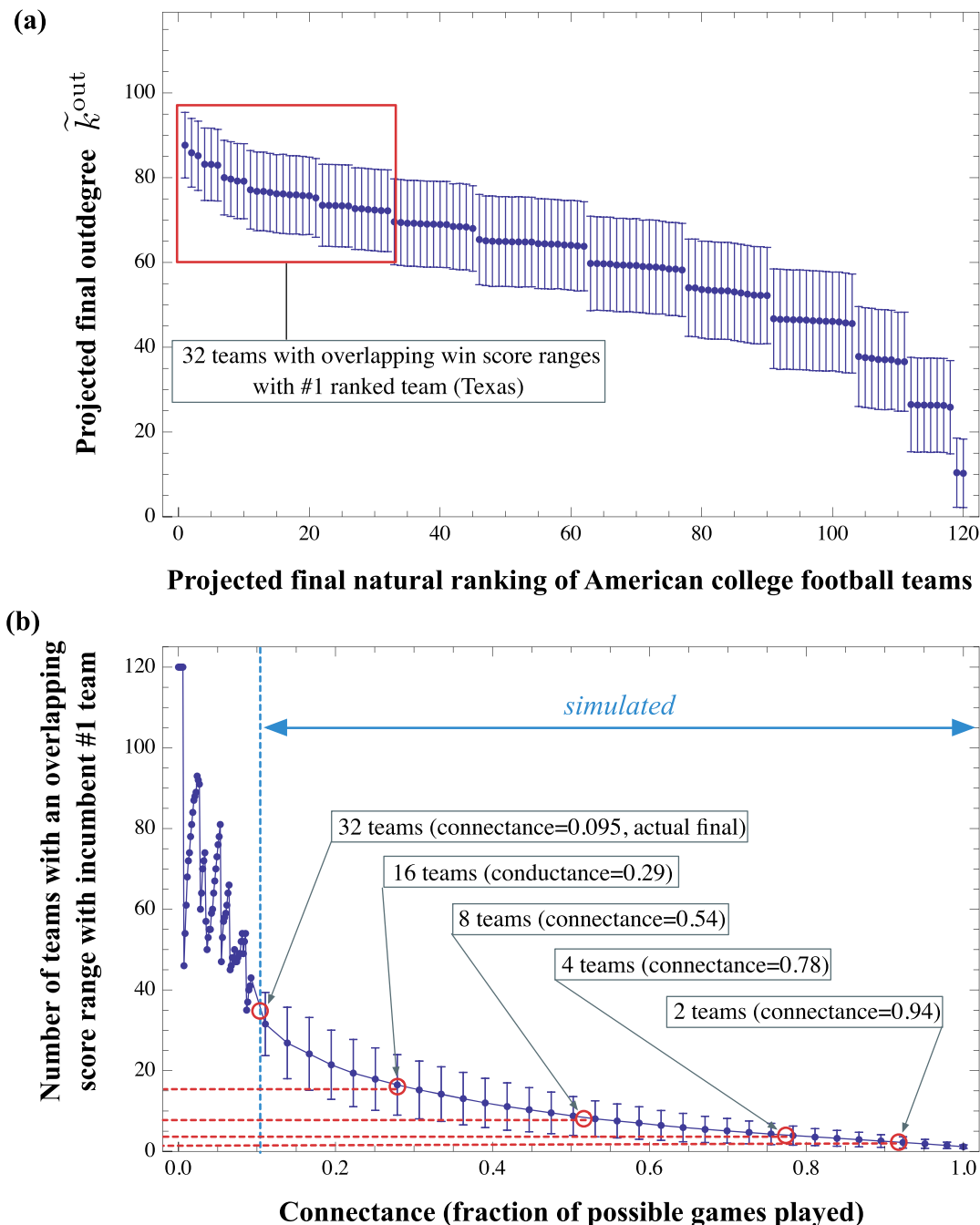


Figure 2 | (a) The calculated projected final outdegree ranges $\left\{ \tilde{k}^{\text{out}} \pm \left(\Delta^2 \tilde{k}^{\text{out}} \right)^{1/2} \right\}$ of American universities that competed in the 2010 network. While the University of Texas (farthest left) has the highest expected \tilde{k}^{out} , its uncertainty $\left(\Delta^2 \tilde{k}^{\text{out}} \right)^{1/2}$ is sufficiently large that its range overlaps with those of 31 other teams, indicating the current playoff system involving two or four top teams is insufficient. (b) We can then estimate the appropriate size of playoff participants by counting the number of teams with overlapping score ranges with the then-highest ranking team as a function of the connectance. The data points indicate the actual numbers of overlapping teams up until the actual final connectance (0.095, blue dotted line), and the estimated numbers past it averaged over 1 000 random simulations. Also indicated are four points at which the estimated mean numbers are 16, 8, 4, and 2; the proposed four-team playoff system would have required a connectance of 0.78 for this season.

same \tilde{k}^{out}) when the first 270 games (connectance = 0.71) were played.

Comparison with Elo and Win-Loss Differential Methods

Of many ranking methods for competitions ELO is one of the best known, adopted by the World Chess Federation (FIDE) and online service providers including Yahoo! Games. We now briefly compare, for illustrative purposes, our method and the ELO devised by physi-

cist Arpad Elo²¹. While there exist several variations of ELO we consider the most basic version. For reference we also compare the simplest method of win-loss differential, i.e. $k^{\text{out}} - k^{\text{in}}$.

In ELO each player is assigned a rating value R . When two teams i and j with ratings R_i and R_j play a game, the probability of i winning plus half the probability of drawing is posited as

$$p_i^E = \frac{1}{1 + 10^{(-R_i + R_j)/400}} = 1 - p_j^E. \quad (16)$$

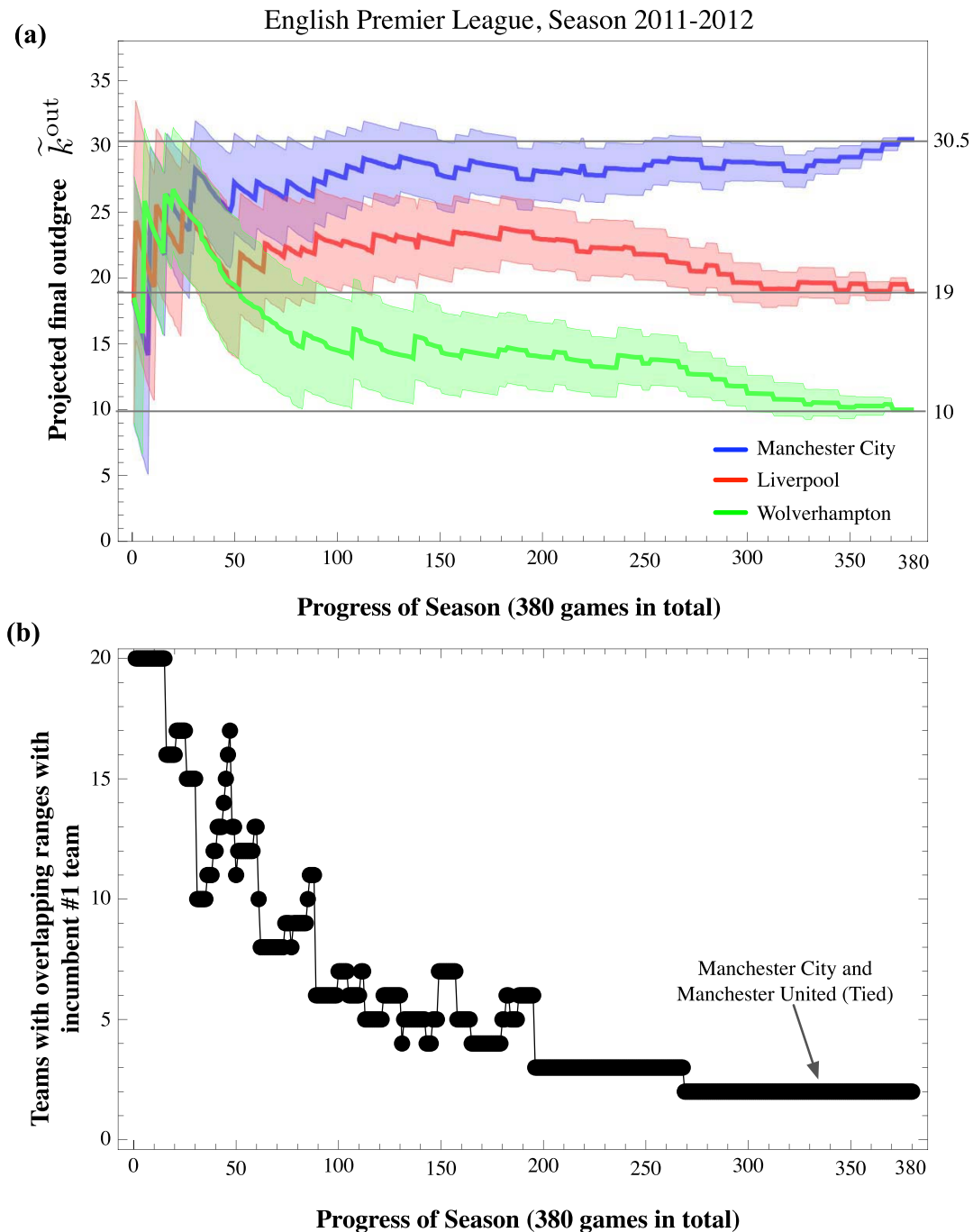


Figure 3 | (a) The changes in \tilde{k}^{out} of three EPL (English Premier League) teams in the 2011–2012 season, representing the top (Manchester City, blue), the middle (Liverpool, red), and the bottom (Wolverhampton, green) tiers. Significant fluctuations and uncertainties dissipate as the season progresses to reveal a clear standing of the teams. (b) The number of teams with overlapping \tilde{k}^{out} ranges with the incumbent #1 team.

A $p_i^E = 0.75$, therefore, could mean 0.5 probability of winning and 0.5 probability of drawing, or 0.7 probability of winning and 0.1 probability of drawing. For each 400 rating point difference against the opponent the ratio p_i^E/p_j^E is multiplied by a factor of ten so that if $R_i = R_j + 400$ we have $p_i^E = \frac{10}{11}$ and $p_j^E = \frac{1}{11}$, and if $R_i = R_j + 800$ we have $p_i^E = \frac{100}{101}$ and $p_j^E = \frac{1}{101}$, and so forth. Elo's original suggestion was that a difference of 200 in ratings points mean that the stronger player has an $p_i^E = 0.75$, as in Eq. (15), $\frac{1}{1 + 10^{-200/400}} = 0.760$. After the game the rating R of i is updated as $R'_i = R_i + 32(\sigma_{ji} - p_i^E)$ and

similarly for R_j , where σ_{ji} is the actual outcome (1 for a win for i , 0.5 for a draw, and 0 for a loss). The initial ELO rating is 1400.

In Fig. 4 we compare the prediction accuracies of the three methods – ours (triangle), ELO (square), and the simplest win-loss scheme (circle) – for American college football. We award each method 1 point for a correct prediction (the team with a higher pre-game rating wins), and 0 point for a wrong prediction. We treat the case of “indeterminate prediction” (two teams tied in pre-game ratings) in two separate ways: First, we include it as a half-correct prediction, awarding 0.5 points to a method; Second, we excluded it (0 points) to consider only the determinate cases as the correct ones. In the first case our Bayesian method and ELO both earn 452 points and win-loss earns 446 points respectively for prediction accuracies of 0.666

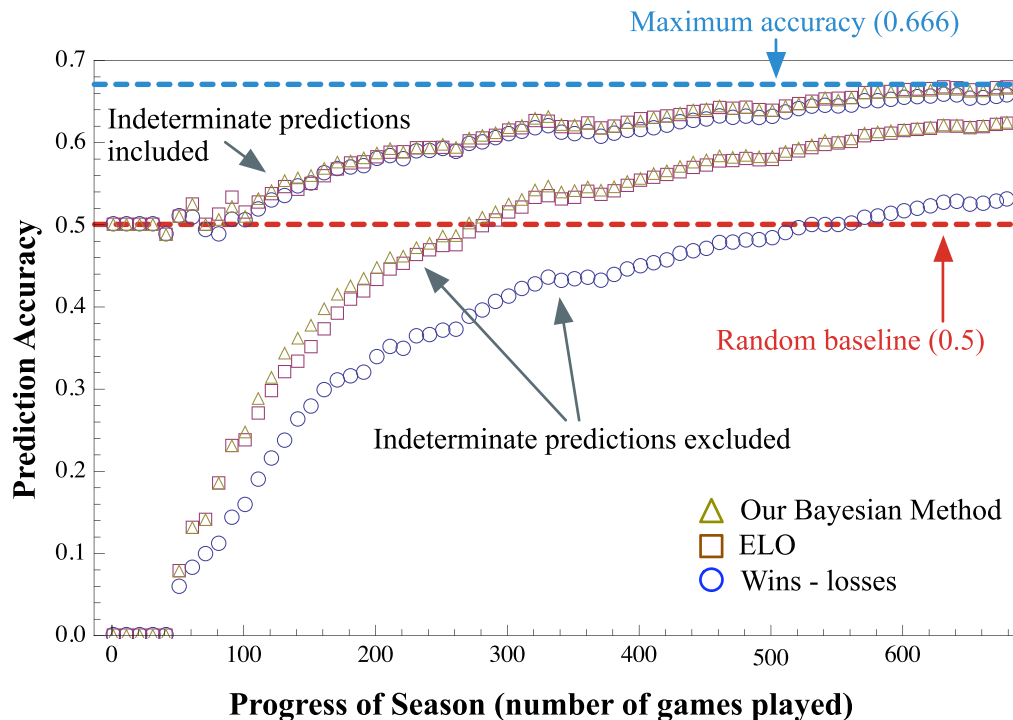


Figure 4 | The comparison between the prediction accuracies of the three methods: our Bayesian (triangle), ELO (square), and the win-loss differential (circle) as the season is played out in American College Football. The red dotted line indicates the random baseline of 0.5. We consider two scenarios regarding the indeterminate-prediction cases (i.e. the teams have an equal pre-game rating): The upper band of plot points includes them as half-correct ones, and the lower band excludes them. The differences between the bands correspond to the fraction of indeterminate predictions, which is notably larger for the wins-losses scheme than the other methods.

and 0.657. In the second case they earn 424, 423, and 360 points for prediction accuracies of 0.624, 0.623, and 0.530. Since the differences between the two plots represent the fraction of indeterminate predictions, it demonstrates the comparative limitation of the win-loss scheme in differentiating two teams based on the wins-losses scheme. As another angle of comparison, we studied how the ratings difference correlates with the points scored in the game, finding Pearson correlation coefficients of 0.756 for our method and 0.731 for ELO. Which method, then, is preferable? Given the comparable performances we would argue that the ability to estimate errors and the solid theoretical foundations render our method preferable to ELO. There exists, as a matter of fact, a deeper connection between the two methods. We start from Eq. (4) which we rewrite as

$$p_{ji} \equiv \frac{\phi_i}{\phi_i + \phi_j} = \frac{1}{1 + \frac{\phi_j}{\phi_i}} = \frac{1}{1 + 10^{(-R_i + R_j)/400}}, \quad (17)$$

the last form being identical to the foundation of the ELO system, Eq. (15) with a change of variables $R = 400 \log_{10} \phi$. The difference is that our method allows R to be negative as necessary and, furthermore, gives R a Bayes' formula-consistent distribution $\pi(R)$ that can be calculated from $\pi(R)dR = \pi(\phi)d\phi$. It is not true of ELO, reflecting the lack of robust theoretical founding.

Discussion

In this report we studied the concept of natural ranking, also called the Condorcet method, in competition networks. While straightforward and intuitive, natural ranking can only be applied exactly to complete networks that are rare in the real world, prompting us to propose an analytical framework for inferring the final natural ranking using Bayes' formula. We formulated a single-variable strength parameter model with exact update rules as new information (wins and losses) is gained. Bayesian inference is fundamentally distri-

bution-based, meaning that it produces not one specific value of a variable but a range of values. This allowed us to estimate not only the mean expectation of the final outdegrees of teams in the network but their uncertainties in variance, enabling us to answer important questions of practical value, such as the sufficient size of a playoff system, and the convergence speed to the final natural ranking.

We envision a couple of future research directions based on the work presented here, one theoretical and the other practical. First, it would render ranking methods — ours included — more useful to understand more analytically how their performances relate to the system parameters such as its size and connectance, although they have been studied here to some extent through various means (calculations and simulations). Second, a further improvement of our method to deal with the aspects of competition networks that were not discussed in this paper would be welcome; for instance, our current model uses a node's past performance as the sole basis for estimating its current strength (fitness), while it can be affected by many other factors. The true distributions of node strengths is deeply related to any ranking method, as one can alternatively view a ranking method as a way to uncover the hidden, true strengths of the nodes. Thus a better estimation technique can lead to better performance and prediction accuracy. Giving more weight or importance to more recent records, for example, may be useful. We hope that such efforts will prove useful for understanding many competition networks better.

1. Newman, M. E. J. The structure and function of complex networks. *SIAM Rev.* **45**, 167–256 (2003).
2. Barabási, A.-L. Taming complexity. *Nat. Phys.* **1**, 68–70 (2005).
3. Barabási, A.-L. & Albert, R. Emergence of scaling in random networks. *Science* **286**, 509–512 (1999).
4. Hagele, G. & Puskelsheim, F. Llull's writings on electoral systems. *Stud. Lull.* **41**, 3–39 (2001).
5. Newman, M. E. J. The physics of networks. *Phys. Tod.* **61**, 33–38 (2008).



6. Cho, Y. S., Hwang, S., Herrmann, H. J. & Kahng, B. Avoiding a spanning cluster in percolation models. *Science* **339**, 1185–1187 (2013).
7. Freeman, L. C. Centrality in social networks: I. Conceptual clarification. *Soc. Net.* **1**, 215–239 (1979).
8. Newman, M. *Networks: An Introduction* (Oxford University Press, Inc., New York, NY, USA, 2010).
9. Page, L., Brin, S., Motwani, R. & Winograd, T. The Pagerank citation ranking: bringing order to the web. *Tech. Rep.*, Stanford University (1998).
10. Lott, D. F. Dominance relations and breeding rate in mature male American bison. *Tierpsychologie* **49**, 418–432 (1979).
11. Balinski, M. & Laraki, R. *Majority Judgment: Measuring, Ranking, and Electing* (The MIT Press, Cambridge, MA, USA, 2011).
12. Williams, R. J. & Martinez, N. D. Simple rules yield complex food webs. *Nature* **404**, 180–183 (2000).
13. MacKay, D. J. *Information Theory, Inference, and Learning Algorithms* (Cambridge University Press, 2003).
14. Jaynes, E. T. *Probability Theory: The Logic of Science* (Cambridge University Press, 2003).
15. Jeffreys, H. An invariant form for the prior probability in estimation problems. *Proc. Roy. Soc. Lon.* **186**, 453–461 (1946).
16. Park, J. & Newman, M. E. J. Statistical mechanics of networks. *Phys. Rev. E* **70**, 066117 (2004).
17. Park, J. Diagrammatic perturbation methods in networks and sports ranking combinatorics. *J. Stat. Mech.* **2010**, P04006 (2010).
18. Stefani, R. T. Survey of the major world sports rating systems. *J. Appl. Stat.* **24**, 635–646 (1997).
19. Callagan, T., Mucha, P. J. & Porter, M. A. The bowl championship series: A mathematical review. *Not. Amer. Math. Soc.* **51**, 887–893 (2004).
20. Dunnavant, K. *The Fifty-Year Seduction* (Thomas Dunne Books, New York, 2004).
21. Elo, A. E. *The Rating of Chess Players, Past and Present* (Arco, New York, 1978).

Acknowledgments

The authors thank Thilo Gross and Naoki Masuda for useful discussions, and Seung-Kyu Shin for assistance with data curation. This work was supported by the National Research Foundation of Korea funded by the Korean government (NRF-20100004910 and NRF-2013S1A3A2055285), Korea Advanced Institute of Science & Technology, Kyung Hee University (Grant KHU-201020100116), BK21 Plus Program for Content Science, and the IT R&D program of MSIP/KEIT [10045459].

Author contributions

J.P. and S.-H.Y. wrote the main manuscript text, and J.P. prepared all figures. All authors reviewed the manuscript.

Additional information

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Park, J. & Yook, S.-H. Bayesian Inference of Natural Rankings in Incomplete Competition Networks. *Sci. Rep.* **4**, 6212; DOI:10.1038/srep06212 (2014).



This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder in order to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-sa/4.0/>